CLustre: semi-automated lineament clustering for palaeo-glacial reconstruction

Mike J Smith[a], Niels Anders[b] and Saskia Keesstra[c]

[a]School of Geography, Geology and Environment, Kingston University, KT1 2EE, UK. michael.smith@kingston.ac.uk. 44 207099 2817
[b]Environmental Sciences, Wageningen University, Droevendaalsesteeg 4, 6708PB Wageningen, Netherlands. niels.anders@wur.nl. 31 317 486617
[c]Environmental Sciences, Wageningen University, Droevendaalsesteeg 4, 6708PB Wageningen, Netherlands. saskia.keesstra@wur.nl. 31 317 482877

Abstract

Datasets containing large numbers (>10,000) of glacial lineaments are increasingly being mapped from remotely sensed data in order to develop a palaeo-glacial reconstruction or ”inversion”. The palimpsest landscape presents a complex record of past ice flow and deconstructing this information into a logical history is an involved task. One stage in this process requires the identification of sets of genetically linked lineaments that can form the basis of a reconstruction.

This paper presents a semi-automated algorithm, CLustre, for lineament clustering that uses a locally adaptive, region growing, methodology. After outlining the algorithm, it is tested on synthetic datasets that simulate parallel and orthogonal cross-cutting lineaments, encompassing 1,500 separate classifications. Results show robust classification in most scenarios, although parallel overlap of lineaments can cause false positive classification unless there are differences in lineament length. Case studies for Dubawnt Lake and Victoria Island, Canada, are presented and compared to existing datasets. For Dubawnt Lake 9 out of 14 classifications directly match incorporating 89% of lineaments. For Victoria Island 57 out of 58 classifications directly match incorporating 95% of lineaments. Differences are related to small numbers of unclassified lineaments and parallel cross-cutting lineaments that are of a similar length.

CLustre enables the automated, repeatable, assignment of lineaments to flow sets using defined user criteria. This is important as qualitative visual interpretation may introduce bias, potentially weakening the testability of palaeo-glacial reconstructions.

In addition, once classified, summary statistics of lineament clusters can be calculated and subsequently used during the reconstruction process.

Keywords: glacial; inversion, palaeo; model; automated; lineament; drumlin

Introduction

Palaeo-glacial landforms are the result of past processes having acted upon the environment; their location, distribution and composition “encode” information concerning the environmental conditions during their formation. By recording at least their location and morphology (e.g. Clark and Meehan, 2001) it is possible to “decode” this information and, making certain assumptions about processes and boundary conditions, infer the extent and dynamics of former ice masses (Kleman and Borgström, 1996; Kleman et al, 2006). A palaeo-glacial reconstruction performed using a geomorphological inversion allows researchers to gain insight into the operation of ice sheets (e.g. Stokes et al, 2009) which can be used to develop a better understanding of the climatic system (e.g. Boulton and Clark, 1990a, b; Boulton et al, 2001; Greenwood and Clark, 2009). The interaction of an ice-mass with the underlying landscape creates a variety of erosional and depositional landforms, such as drumlins, eskers, meltwater channels and moraines.

By using palaeo landform evidence as a proxy (e.g. landform size, shape, length, orientation), a reconstruction can be used to calibrate and test existing climate models helping to parameterise the integration of feedback mechanisms from oceanic and cryospheric systems (e.g. Hubbard et al, 2009).

The process for performing a palaeo-glacial reconstruction has developed over a substantial period of time, dating back to the first mapping of large suites of landforms in the Nineteenth century and their interpretation (e.g. Close, 1867a, b; Kilroe, 1888; Wright, 1912, 1937; Charlesworth, 1924, 1939). Subsequently an understanding of process-landform assemblages developed, as well as the manner in which palimpsest evidence was interpreted (e.g. Sugden and John, 1976). A significant advance was the understanding that multiple ice flow phases can be recorded in the palimpsest landscape (Rose and Letzer, 1977; Boulton and Clark, 1990a) through either the remoulding of existing landforms or the superimposition of a newer landform on top – termed cross-cutting (Clark, 1993). Not only did this recognise that the contemporary land surface could be a heterogenous record of past events, but that in these instances the relative age of landforms could be ascertained. However studies generally remained at a local scale (e.g. Rose, 1989) due to the extensive time required to map large regions in a consistent manner (e.g. Rose and Smith, 2008). The limited spatial scope of these studies prohibited

landscape scale interpretation and limited the potential to use these data for ice sheet reconstruction.

To improve palaeo-glacial reconstruction for understanding past climate change , three significant methodological improvements can be observed over the last 40 years. Firstly, the widespread availability of moderate resolution satellite imagery enabled the mapping of landforms across large regions consistently and economically (e.g. Punkari, 1980), resulting in the compilation of large databases of landforms (e.g. Stokes et al, 2013). These have incorporated a range of landform evidence including drumlins (or more generically lineaments), moraines, ribbed moraine and meltwater channels (e.g. Dunlop and Clark, 2006; Margold et al, 2011). As sensor spatial resolutions have improved, so it has been possible to map smaller features (e.g. Hughes et al, 2010). Secondly, digital elevation models (DEMs) have enabled the detection and mapping of many landforms, as well as improving the ability to produce quantitative measurements (Smith and Clark, 2005; Clark et al, 2009). Thirdly, developments in the interpretation of palaeo landform evidence (e.g. Kleman and Borgström, 1996; Clark, 1997; Kleman et al, 2006) has led to more informed and better constrained numerical modelling of past ice sheets.

The production of a palaeo-glacial model from landform evidence (Figure 1a) requires: (i) a base dataset (Figure 1b); (ii) landform mapping (Figure 1c); (ii) data reduction (generalisation; Figure 1d); (iii) assignment of relative chronology and (iv) interpretation (e.g. Clark, 1997; Kleman and Borgström, 1996; Smith and Knight, 2011). More specifically satellite imagery or DEMs are used as data sources to manually map individual glacial landforms. For regions containing large suites of landforms, which may involve tens of thousands of individual features, this is a slow process. A data reduction stage is required (Smith and Knight, 2011) whereby qualitative, operator-based, visual heuristics are used to correlate regions of internally homogenous features to create *flow patterns*. Finally, these can be merged over larger regions to form coherent units called *flow sets* (Clark, 1997) or *fans* (Kleman and Borgström, 1996); at this stage they are assigned a relative chronology and can then be used as a basis for an interpretation of the local and regional ice flow dynamics.

The focus of this paper is upon *automating* the second and third stages; namely taking individual landforms to create “spatially coherent” flow patterns. This will enable the repeatable identification of flow patterns, a stage that currently requires

qualitative interpretation. Here we present a semi-automated algorithm for the clustering of landforms, specifically lineaments (CLustre). The objective of this paper is to show the usability of the algorithm to assess lineaments and extract flow patterns over large areas automatically, thereby allowing the user to subsequently interpret flow sets. We begin by proposing criteria for defining “spatially coherent” before outlining the algorithm and implementation. The algorithm is first tested on a synthetic dataset, before application to two contrasting case studies. One significant benefit is the ability to produce summary statistics of individual flow sets, something that is time consuming to do manually, particularly where flow sets are overlapping.

Method

Spatial Coherence

The production of a glacial inversion model requires the reduction, or grouping, of large numbers of individually mapped landforms to create “spatially coherent” flow patterns (Figure 1). Landform mapping is usually performed within a geographic information system (GIS) and requires dimensional abstraction to either points, lines or polygons (Smith, 2011). Mapping has typically been performed using lines as these are able to represent linear features (marking their ridge crest), however the availability of high spatial resolution remotely sensed data has seen a move towards the use of polygons through mapping the bounding concave break-of-slope. For the lineament clustering presented in this paper it is necessary to use line geometry with a start and end connected by a single line. This can be easily be derived from polygons within a GIS. In addition to lineaments, moraines, ribbed moraine and meltwater channels have all been used as inputs to inversion models (Kleman and Borgström, 1996), however lineaments are by far the most numerous.

The assessment of “spatial coherence” needed for the identification of flow patterns is a qualitative process that involves making a visual judgement as to the continuity of patterns identified in the mapping (Figure 2; Clark, 1999). This process is based upon the use of heuristics, both implicit and explicit, to reduce large numbers of individual observations to fewer, representative, flow patterns (and then flow sets). The heuristics used to perform this procedure can be classified as two types: (i) spatial and (ii) contextual.

Spatial heuristics are based upon the two dimensional spatial organisation of individual landforms relative to one another. Interpretation of spatial organisation is dependent upon the method used to make observations and subsequent metrics to describe them. For lineaments mapped as lines (Figure 1) we can define *orientation* and *length* for individual features and *density* for groups of features. Morphometric similarity (orientation, length and density) can then be used as a basis for grouping lineaments (Figure 2).

Contextual heuristics use expert/operator knowledge *about* features to assign membership to flow patterns. Figure 3 demonstrates examples of two scenarios where membership to a flow pattern would not occur using just spatial heuristics (see the discussion for further detail).

When interpreters make qualitative judgements, they combine spatial and contextual heuristics to assess pattern *continuity* and subsequently reduce large numbers of observations to fewer, more representative, ones. Our approach to semi-automate this process is to mimic the process of grouping, based upon orientation, length and density, and attempt to build-in processes to aid contextual grouping or allow an interpreter to assist the process. Computationally this is known as *clustering* – assigning lineaments to groups, based upon their attributes. This brings three primary benefits: (1) semi-automatic identification of flow patterns; (2) objective and interactive exploration of potential flow patterns; and (3) automatic grouping of lineaments for statistical analysis.

Algorithm Development

The method presented here clusters digitised lineaments and is based upon the initial work of Smith (2003). CLustre uses a vector based region-growing methodology in order to mimic the visual assessment of homogenous regions, whilst allowing for growth taking into account gradual spatial variations. It begins with an initial operator-selected lineament (or seed lineament), and groups neighbouring lineaments according to their distance (from the seed lineament), orientation and length (Figure 4). From the newly grouped lineaments, two new seeds are selected to continue the clustering procedure, iterating over until pre-defined thresholds, discussed below, are reached. CLustre is written in the Python programming language, and uses the open-source GDAL/ OGR library (http://www.gdal.org/). Interactive examination of the input and output files was undertaken using QGIS

(http://www.qgis.org). CLustre is freely available for download (https://github.com/niels-anders/clustre).

Prior to processing lineament data, the interpreter is required to input user defined thresholds (Table 1). CLustre then requires a single input file comprised of an Esri polyline shapefile containing lineaments represented as single lines (two nodes, linked by an arc) recorded in Cartesian coordinates; the script ignores lineaments with more than 2 nodes. Prior to the iterations, the start and end node x,y coordinates are used to calculate the length and orientation of each lineament, with orientation corrected to fall within the 0-180° range. The algorithm then proceeds in the following manner (Figure 4):

1. Begin at the seed lineament (*s*);
2. Select all lineaments that are closer to *s* than the given threshold (*d*) ;
3. Compare the **orientation** of *s* with the selected lineaments; sub-select those that do not differ by more than the given threshold *α*;
4. Compare the **length** of *s* with the sub-selection and select those that do not differ more than the given threshold *l*;
5. Add the current selection to the cluster;
6. Two new seeds are assigned from the selection created in (5). The first new seed is the lineament that is located farthest from *s*, this becomes s_1;
7. Calculate the position of s_1 relative to *s*;
8. Select the second new seed (s_2) as farthest away from *s* in the opposite direction;
9. Both new seeds are added to a list of seeds that need to be processed. From there, the procedure iterates from (2) until all seeds are processed. Previously used seeds and classified lineaments cannot be reused.

The use of farthest apart secondary seeds is designed to reduce the number of distance calculations whilst allowing expansion of the algorithm into all parts of the study area. The algorithm then simply operates by testing all lineaments for spacing (i.e. distance to its nearest neighbour), orientation conformity and length conformity. Output from CLustre is the original shapefile with five additional parameters: *id*, *label* (flow pattern number), *seed* (whether it was a seed), *order* (order of classification) and *source* (the source lineament for classification). Unclassified lineaments have *label* values set to zero which enables subsequent processing of a previously classified dataset. This is an important aspect of developing a set of flow patterns as

it allows the end user to iteratively process and classify lineaments, developing groups that can then be used to produce flow sets.

Synthetic Data

In order to assess the operation of CLustre, three simulated scenarios (Figure 5) were produced which depict straight, curving and divergent (or convergent) lineaments. This spatial organisation is commonly found in real datasets and was a requirement of the testing. The scenarios incorporate the use of both one and two flow-sets to contrast a simple glaciological scenario with a more complex cross-cutting one, with the latter tested for variations in lineament length and orientation.

The synthetic lineaments were generated using the following procedure:

1. Define spacing (d_s), orientation (α_s) and length (l_s) for the synthetic lineaments;
2. Define the **variance** for spacing (d_{var}), orientation (α_{var}) and length (l_{var}) of the synthetic lineaments;
3. Define the size (A_s) of the synthetic area ;
4. Calculate synthetic lineament x,y location based upon A_s and d_s to create the "straight" scenario;
    4a. Progressively rotate each row (or column) of lineaments to create a "curving" scenario - greater amounts of rotation increase the amount of curvature;
    4b. As per 4a, a curving pattern is generated and then mirrored to create a "divergent" scenario;
5. Randomly add d_{var}, α_{var}, l_{var} to each lineament.

The synthetic area was fixed at 150x150 units with a total of 100 lineaments. Step (4a) allows the generation of a curvature to the lineaments and (4b) modifies this to create a divergent pattern. For testing, d_s, α_s and l_s values for the first flow-set (FS1) were set at 15, 2 and 5 respectively, with values of 15, 92 and 5 for the second flow-set (FS2). Variance was also applied to spacing (d_{var}=1), orientation (α_{var}=2)_and length (α_{var}=0.5) in order to better simulate a real-world dataset.

Figure 5 shows the 15 synthetic datasets that were generated for testing from the three scenarios. For the straight (Figure 5a), curving (Figure 5f) and divergent (Figure

5k) scenarios our expectation is that CLustre should correctly identify the flow patterns. In these cases, variations in length of lineament datasets are not relevant as it is a characteristic related to scale and is accommodated by manually selecting a high length threshold that allows the classification to operate effectively.

Figure 5 also shows more complex two flow-set examples that test the CLustre algorithm for variations in lineament orientation (using end-members) and length:

1. Straight Lineaments: end-member examples with lineaments aligned orthogonal and parallel to one another. For orthogonal alignment we expect CLustre to successfully classify lineaments based upon variations in orientation (Figure 5b) and length (Figure 5d). For parallel alignment it is not possible to classify based upon orientation (Figure 5c), only length (Figure 5e).
2. Curved Lineaments: end-member examples with lineaments aligned orthogonal and parallel to one another. As for straight lineaments, orthogonal alignment should successfully classify for orientation (Figure 5g) and length (Figure 5i). For parallel alignment it is only possible to classify based upon length (Figure 5j).
3. Converging/Diverging Lineaments: end-member examples with lineaments aligned orthogonal and parallel to one another. As for straight lineaments, orthogonal alignment should successfully classify for orientation (Figure 5l), and length (Figure 5n). For parallel alignment it is only possible to classify based upon length (Figure 5o).

In order to remove any bias associated with the selection of the seed lineament, each of the 15 synthetic datasets was classified iteratively using each of the 100 lineaments as a seed – in total there were 1500 classifications performed. A classification was considered successful if it classified all FS1 lineaments and did not classify any FS2 lineaments. User thresholds were set for distance

$$d = \sqrt{2{d_s}^2} + \sqrt{2{d_{var}}^2} + 0.001 \approx 24.04$$

as well as for orientation (α=6) and length (l=2). This was designed to maximise inclusivity of the classification - the distance radius was set slightly larger than the maximum (diagonal) distance between two lineaments, with orientation and length thresholds larger than the variation between directly neighbouring lineaments of a single flow pattern.

Figure 6 and Table 2 present results for the classification of the synthetic lineament data. Only seeds for FS1 were used, which means that a "correct" classification is where all FS1 lineaments are classified and all FS2 lineaments are *not* classified. Figure 6a and 6b show results for the straight flow sets; for FS1, nearly all 100 simulations classified the 100 lineaments correctly (Table 2). No lineaments from FS2 were misclassified. In a small number of classifications, not all FS1 lineaments are classified – this shows the potential impact of the seed lineament and, in this case, the effect of lineaments at the edges causing the classification to partially fail. Figure 6c shows that, as expected, nearly all FS2 lineaments were misclassified when the flow sets were aligned in parallel. Similar results were achieved for the curved flow sets (Figure 6f, 6g and 6h). For the convergent lineaments a more complex picture emerges – Figure 6k shows that for ~10% of classifications only ~50% of FS1 lineaments were classified. For the orthogonal flow sets (Figure 6l) a highly complex picture is apparent and whilst for ~80% of classifications nearly all FS1 lineaments were correctly classified, in ~80% of classifications ~60% of the FS2 lineaments were incorrectly classified as FS1. For the parallel aligned flow sets (Figure 6m) lineaments from FS1 are correctly identified in nearly all classifications, although nearly all FS2 lineaments were misclassified.

When the length of lineaments is varied between both flow sets, the ability to discriminate between flow sets is dramatically improved. As before, for the straight (Figure 6d) and curved (Figure 6i) orthogonal scenarios nearly all 100 classifications classified the 100 lineaments correctly for FS1, with no lineaments from FS2 misclassified. With the introduction of a difference in length, the straight (Figure 6e) and curved (Figure 6j) parallel scenarios are now correctly classified. In the more complex divergent scenarios, nearly all the orthogonal classifications (Figure 6n) correctly identified FS1 with no misclassification of FS2, with a similar result achieved for the parallel scenario (Figure 6o).

Discussion

Results clearly show the efficacy of CLustre in classifying lineaments into individual flow sets. Whilst *spatial coherence* can potentially be assessed on the basis of orientation, length and density, density cannot be used as it requires *a priori* knowledge of the flow set extent. The success of CLustre comes from replicating the visual process that interpreters use, but doing so more efficiently. That is, it is based upon proximity, then assessing similarity of orientation and length – as a “region growing” clustering technique it can locally adapt, within thresholds, to natural

variations within the data. The results also demonstrate the importance of using both length and orientation in a successful classification. Where there is one flow set, or flow sets that cross-cut one another obliquely, it is relatively simple to classify lineaments that are straight or curving. It is when lineaments from different flow sets are aligned parallel to one another that classification becomes more difficult. And whilst Table 2 shows overall accuracy is high, user accuracy (Congalton and Green, 1999) is significantly reduced by the misclassification (i.e. false positives) of FS2 lineaments (e.g. 50.2% for straight flow sets). By introducing the second constraint of lineament length it is possible to successfully classify parallel flow sets for both straight and curving lineaments, assuming a consistent difference in length is present between the two flow sets.

Divergent lineaments are more complex as they represent a single, coherent, spatial pattern that combines together both straight and curving geometric elements that are mirrored around a centreline. This means that, even for orthogonal flow sets, there remain individual lineaments that are aligned parallel to one another. This makes classification more complex with a large number of misclassifications for the orthogonal flow set and almost complete misclassification for the parallel flow set. With the introduction of length as a further constraint, classification is successful.

The results also show that CLustre is not particularly sensitive to the seed lineament used to begin the classification – this is an important finding as the user can be confident in the classification result. There are clearly instances where the selection of some lineaments as a seed can decrease accuracy (Table 2). For example when (i) located on the edge of a study area or (ii) parallel to a lineament from another flow set. The latter is particularly important when there are curving or divergent flow sets. However Figure 6 demonstrates that most simulations converge on a good solution. As CLustre is similar in general method to the common k-means classifier (Lillesand et al, 2008; Schowengerdt, 2006) used with raster datasets, it is expected that where there are well defined, separable, flow patterns the seed lineament will be less sensitive (Congalton and Green, 1999). With testing completed on the synthetic data set, CLustre was subsequently applied to real world data sets in order to test its applicability to two contrasting examples.

Case Studies

Dubawnt Lake

The first case study comprises 11,825 lineaments, covering ~100,000 km^2 mapped from the Dubawnt Lake region, on the border between Nunavut Territory and Northwest Territories in the Canadian Arctic (Stokes, 2000; Stokes and Clark, 2003). This area is comprised of mega-scale glacial lineations (MSGL) which are thought to be indicative of fast ice flow over relatively short periods of time; these are interpreted as representative of a palaeo-ice stream that drained the interior of the Laurentide Ice Sheet during the last glaciation (Stokes and Clark, 2003). The landforms are elongate and highly attenuated – they are longer than drumlins, typically 10-100 km, yet only 300-1300 m wide (Stokes et al, 2013) and densely packed.

The Dubawnt Lake data were selected as providing a simple case study with one *primary* flow pattern defined by the highly attenuated lineations in the form of MSGL and drumlins. Additionally, there are a smaller number of other lineaments of varying orientations and spacings that provide contrast to the main flow pattern (Figure 7a). Note that the straight edges (top of figure) mark the Landsat Multi-Spectral Scanner scenes that were used to map lineaments in the original work (Stokes, 2000).

The workflow initially comprised of visually identifying the primary flow pattern and selecting a seed lineament. Based upon the results for the synthetic data, this should not be near an "edge" or parallel to another flow pattern. A lineament approximately in the centre of the flow pattern and of "average" length and orientation was selected (Table 3). With many MSGLs 2-3 km in length and with high attenuation and density, a distance threshold (d) of 10 km was selected. And, given the high attenuation, a relatively small orientation threshold (α) of 2º was used. Finally a relatively large length threshold (i) of 3 km was chosen given the wide range of MSGL lengths (Figure 7a). Inspection of the classification suggests that the thresholds have isolated this flow pattern, although there are some lineaments that remain excluded due to threshold exceedance. Once a flow pattern was classified it was then hidden, leaving the remaining lineaments visible; each flow pattern was then iteratively classified in the same manner

An interactive approach for establishing threshold values (Table 3) was taken, reviewing the classification results using QGIS. Given that CLustre is not sensitive to the seed lineament, if a classification produced unexpected results the reason for this was explored further. This often resulted in adjusting the threshold values and then re-running CLustre. Figure 7b shows the application of threshold values from flow

pattern 1 to flow pattern 2. This is successfully grouped 93 lineaments but a small adjustment of thresholds produced a better classification with 103 (Figure 7c). Similarly flow pattern 6 was classified using the initial thresholds (Figure 7d; 128 lineaments), however visual inspection of this pattern indicates a low density of short lineaments, with some clustering. Threshold values were adjusted to a larger distance (35 km) and orientation (5º) resulting in a significantly better classification (Figure 7e; 1015 lineaments).

Once classified, it is possible to report quantitative measures of spatial coherence for each flow pattern (Table 3); these include the number and density of lineaments, and summary statistics for lineament length. Density was calculated by dividing the number of lineaments by the area of the convex hull (smallest bounding area) for each flow pattern. For illustration, Figure 8 shows histograms of length for each flow pattern, demonstrating clear differences between them. FP1 has, by far, the most lineaments with many MSGLs, interspersed with drumlins, producing a relatively large mean and wide range. Whilst FP2 has a greater mean and standard deviation (SD), it is a relatively small flow pattern. FP3 and 4 are similar to FP2, although with fewer longer lineaments. FP5, 6, 7 and 8 generally have the shortest lengths and lowest SD, with FP6 and 7 having the lowest densities.

Figure 9 shows the outcome of the classification process with a total of eight flow patterns, with 1,285 (11%) lineaments unclassified (termed residuals; black). Flow pattern 1 (FP1; red) forms the main classification, however it is clear that a small number of residuals have been excluded due to variations in length (of MSGLs) and orientation. FP2 (dark grey) and FP4 (brown) form two low angular cross-cutting classifications that have been separated, however residuals from both flow patterns have been excluded on the basis of length and orientation. FP3 (blue) forms a distinct classification, although a number of small groups of residuals could be incorporated into this and have been excluded due to variations in spacing. FP5 (pink; partially on Figure 8) forms a separate group of low density, short, lineaments. FP6 (green) is comprised of small groups of short, moderately dense, lineaments. The large spacings between groups required a significant increase in this threshold. FP7 (sky blue) form a small coherent group on the southern margin and cross-cut FP8 (light grey). The remaining notable area of residuals is on the western margin of the study area; this is a complex area of converging ice flow for FP1, whilst also having numerous lineaments of varying orientations. Whilst some could be

incorporated into FP1, their small number and varying deviations in orientation make them anomalous and open to interpretation; they therefore remain unclassified.

Stokes (2000) identifies and reports descriptive statistics for 14 flow patterns. Of these 9 are matched in this study and show very similar descriptive statistics for lineament morphology (Table 3), representing 89% of all lineaments. Unmatched flow patterns of Stokes are 2 (38), 5 (45), 8 (59), 9 (142) and 13 (95) (totals in brackets). Confidence with the definition of a flow pattern comes from the number of lineaments, areal coverage and density; as these increase there is greater certainty in their classification and so their use as palaeo-glaciological indicators. Within this context Stokes' (2000) flow patterns 2, 5 and 8 have low numbers, whilst 9 and 13 are also of low density. Whilst these residual lineaments could be classified, they have deliberately been excluded due to uncertainty over their grouping.

Victoria Island

The second case study, covering Victoria Island, Canada (Storrar and Stokes, 2007; Stokes et al, 2009), is much larger than the Dubawnt Lake case study in terms of both area (~225,000 km^2) and number of lineaments (54,468). This large landmass also spans the border between Nunavut Territory and Northwest Territories in the Canadian Arctic Archipelago and has a rich array of well-preserved glacial landforms, particularly lineaments. The glacial history is complex, with lineaments comprising straight, curved and converging/diverging patterns that cross-cut one another (Stokes et al, 2009). They are often discontinuous with “patches” of lineaments interspersed with regions free from glacial bedforms. As a result it is a complex task to develop a palaeo-glacial reconstruction for this area due to the partial, intersecting, palimpsest landscapes. It is therefore one of the most challenging regions that can be used for an automated routine.

A similar procedure was used to classify the lineaments for case study 2 as was used for case study 1, however in this instance there are a much greater number of lineaments, covering a larger area and segmented into more flow patterns of varying orientation, length and density. In total 57 flow patterns were identified (Figure 9), leaving 3,038 (5%) residual lineaments.

For comparison, results from Stokes et al (2009) were used. Whilst both Stokes (2000) and Stokes et al (2009) provide a schematic outline of each flow pattern, the former produced flow pattern morphometric statistics (for far fewer groups) which

made direct comparison straightforward. For the latter it was necessary to work from the schematic diagrams (their Figure 6). Stokes et al (2009) identify 71 flow patterns, of these 57 are directly matched or partially matched (see discussion below), 1 is matched by residuals, 6 are not included in their reconstruction and 7 appear not to be based on underlying lineaments. Table 4 illustrates the flow pattern statistics, showing a high lineament density for many flow patterns.

Review of the residuals shows four different types of lineaments that are unclassified:

(1) N-S MSGLs: on the eastern side of the island there are a moderate number of MSGLs that run N-S. These cross-cut FP 2, 7 and 6, and are aligned with 1 and 9. Stokes et al (2009) incorporate the aligned lineaments within those flow patterns, but have not classified the cross-cutting lineaments. It is possible these form one flow pattern but this is open to interpretation.

(2) E-W MSGLs: FP3 is a highly complex region of partially aligned lineaments. A number of MSGLs remain unclassified due to significant differences in length with other lineaments in this flow pattern. Indeed Stokes et al (2009) classify three flow patterns in this area (53-55) based upon cross-cutting relationships and this highlights that unless there is a significant difference in length, it is not possible to differentiate between parallel flow patterns and so this would have to be performed manually.

(3) Small Numbers: a number of smaller "clusters" of lineaments could be further classified, however, as per the first case study, these are comprised of low counts and have therefore been left as residuals (e.g. between FP5 and 6).

(4) Isolated: across the study area there are individual isolated lineaments that have not been incorporated into a flow pattern as they do not meet orientation, size or distance criteria for classification (e.g. between FP5 and 7).

Table 4 also shows matched flow patterns with Stokes et al (2009). Of the 71 flow patterns there are three broad categories of equivalence: (a) a direct match from CLustre to Stokes et al (2009) (e.g. CLustre FP18) (21 flow patterns). (b) a match of multiple CLustre classifications to a single Stokes et al (2009) (e.g. CLustre FP28,29) (9 flow patterns). (c) a match of a single CLustre classification to multiple Stokes et al (2009) classifications (e.g. CLustre FP3 and FP2) (21 flow patterns).

These results show outcomes which are largely expected; there is good direct equivalence between flow patterns, as well as overlapping equivalence of both multiple CLustre and multiple Stokes et al (2009) flow patterns. For (a), FP18 (Figure 10) is a good example of a direct match where a simple flow pattern has the same classification. For (b), the combination of FP26 and 42 reflects the classification of two discrete patterns into a single flow set. This illustrates interpretation by the observer, merging two flow patterns to form the final flow set.

The greatest differences however arise from the last category of equivalence; FP3, as noted above, is a highly complex region of partially aligned lineaments. From visual interpretation, it is apparent that this region has a number of MSGLs, as well as many lineaments with discordant orientations. The creation of flow patterns by Stokes et al (2009) is primarily dependent upon cross-cutting relationships visible on the satellite imagery, augmented with the morphological characteristics noted above. In the absence of this information an automated routine will be unable to differentiate parallel flow patterns.

In contrast, FP2 corresponds to (at least) four Stokes et al (2009) flow patterns (SFP27, 39, 42, 43) and reflects a different classification of lineaments. It is worth noting that Stokes et al (2009) also used a range of other landform indicators, including eskers, moraines and meltwater traces. They also split apart multiple time transgressive deglacial flow sets. SFP27 is the core flow pattern, however SFP42 and 43 show length and orientation concordance and have been grouped together. The SW terminus of SFP27 includes lineaments from FP7, yet CLustre showed significant orientation differences and did not classify them together. In addition the NW terminus of SFP42 does not include the continuation of FP2 which has strong orientation conformity. This is not to state that either flow pattern is "correct", but rather it highlights how objective approaches to forming flow patterns can aid objective interpretation.

Discussion

The two case studies were successful in demonstrating the application of CLustre to existing data and the ability to classify individual flow patterns by iteratively building objective groups of lineaments based upon a consistent methodology. In comparison to a manual method, it is rapid and reduces any bias that may be inherent in any subjective procedure. The process required the operator to check all classifications

and review reasons for particular classification outcomes. This enabled an iterative process where threshold values could be modified based upon the flow patterns that were identified.

It became apparent whilst reviewing classification results from CLustre and comparing them to our own (manual) subjective assessments that we subconsciously used visual biases. Of particular note was the natural grouping of low density (or "short") lineaments; in these instances, subjective assessment placed less emphasis upon orientation variation meaning that increasing dissimilarity of orientation was incorrectly used as a basis for manual classification in these situations.

In addition to the application of the above objective workflow, CLustre was also used to produce quantitative measures of flow patterns that could then be used to support subsequent interpretation and the development of a full palaeo-glacial reconstruction. Descriptive statistics of lineament length and density can be used to demonstrate quantitative differences between flow sets (Table 3 and 4, Figure 8).

There are two potential weaknesses in the results outlined above. Firstly, CLustre calculates the distance between lineament *centroids*; in the instance where a long lineament is surrounded by much shorter lineaments, the lineaments may actually be close together yet the centroids far apart. Whilst this would likely affect MSGLs, no examples were identified in the case studies, possibly due to the ability of CLustre to grow its classification spatially. Secondly, lineament density is calculated using a convex hull; whilst calculating a minimum bounding area, it doesn't represent the area covered by lineaments and so will likely under-estimate density.

Wider Interpretation

Application to Palaeo-Glacial Reconstruction

Once flow sets have been created, a full palaeo-glacial reconstruction requires their mode of emplacement to be ascertained. It is important to note that there are a variety of different glaciological scenarios thought to be able to generate flow sets (Kleman and Borgström, 1996; Clark, 1999; Kleman et al, 2006). These can be categorised as either time transgressive (formed over a period of time) or synchronous (formed at a point in time). Synchronous flow sets exhibit high conformity of orientation and length over small areas, with gradual and systematic changes over larger areas (Clark, 1999). Time transgressive flow sets are formed

during periods of varying regimes of ice flow and consequently display obvious discordancy, with lower conformity of orientation and length, including unsystematic cross-cutting (Clark, 1999). Time transgressive lineaments may form behind a retreating (deglacial) ice margin which would account for this discordancy.

In general, synchronous flow sets are easily identified, even when they are cross-cutting as they have high orientation conformity and gradual changes in geometry. However, there are two scenarios which are more complex and so more difficult to identify:

1. Low Angular Cross-cutting (Figure 3a) – flow sets which cross-cut at low angles are very difficult to identify, even by manual techniques, as lineaments become sub-parallel and the criteria for their classification non-unique. In the example illustrated, a synchronous flow set has low-angular cross-cutting with another synchronous flow set. They may be distinguishable through differences in orientation, but at the southern margin lineaments may be oriented in the same direction. The iterative nature of CLustre allows the interpreter to review the flow patterns they create, enabling them to identify potential cross-cutting. Whilst length is the only threshold that could be used to differentiate these flow patterns, the operator can at least be aware of the potential for misclassification.

2. Time-Transgressive (Figure 3b) – the diagnostic criteria for time-transgressive flow sets are contrary to all the techniques used to identify synchronous flow sets. There can be cross-cutting (low to medium angular differences) within flow patterns, abrupt changes in morphometry and low orientation conformity. CLustre is particularly helpful here as variations in orientation can be rapidly identified. However interpretation can become complex if there are topographic constraints or if, in addition, there is low-angular cross-cutting.

In both of the above cases further morphological evidence in the form of cross-cutting and related landforms (i.e. eskers, moraines, meltwater traces) can aid in distinguishing between flow sets.

Optimisation and Further Development

CLustre enables the automated classification of individual flow patterns based upon user thresholds and the identification of a seed lineament. Further development could seek to optimise the parameterisation of thresholds and present the outcomes of a

large number of multiple classifications to the interpreter. This would allow a range of scenarios to be reviewed and could be extended to enable a lineament to “belong” to multiple flow patterns, assigning probabilities based upon the input thresholds used in that classification. Making use of cross-cutting information could form a valuable part of this process if the superimposition (or Z-level) of lineaments was recorded.

Further development could also automate the calculation of flow pattern statistics, including the calculation of the convex hull and lineament density.

Conclusions

This paper has introduced a semi-automated algorithm for lineament clustering, CLustre, that uses a GIS-based region-growing, locally adaptive, methodology. Processing requires the user to specify thresholds for lineament orientation, length and spacing in order to follow a rule-based classification procedure for the inclusion of a particular lineament within a flow pattern. A “seed” lineament is specified by the user and this forms the basis from which groups of lineaments can be “grown” into flow patterns. The remaining lineaments remain excluded from flow pattern membership and subsequent classifications, using different seed points, can then be used to identify additional flow patterns.

CLustre was tested using synthetic datasets representative of lineaments formed in common palaeo-glaciological scenarios. These demonstrate that it is not sensitive to the initial seed lineament used and that it produces consistent results. It was then applied to two existing datasets that were produced by manually digitising lineaments depicted on satellite imagery; these represented contrasting simple and complex glaciological scenarios. The data encompassed 11,825 and 54,468 individual lineaments, with a range of lineament orientations, lengths and densities from different glaciological scenarios. CLustre was used to classify 8 and 57 flow patterns respectively that were produced through an iterative process to classify, review and re-classify flow patterns. Summary statistics of lineament length and density were produced and would enable an interpreter to statistically test differences between flow patterns and group them into glaciologically meaningful flow sets.

CLustre enables three significant benefits: (1) the semi-automated procedure is fast; (2) it applies an objective, repeatable methodology and (3) once grouped, summary statistics of flow patterns can be assessed.

Acknowledgements

We would like to thank Chris Stokes for the Victoria Island and Dubawnt Lake lineament data.

References

Boulton GS, Clark CD. 1990. A highly mobile Laurentide Ice Sheet revealed by satellite images of glacial lineations. Nature 346: 813-817

Boulton GS, Clark CD. 1990. The Laurentide Ice Sheet through the last glacial cycle: the topology of drift lineations as a key to the dynamic behaviour of former ice sheets. Transactions of the Royal Society of Edinburgh: Earth Sciences 81: 327-347

Boulton GS, Dongelmans P, Punkari M, Broadgate M. 2001. Palaeoglaciology of an ice sheet through a glacial cycle: the European ice sheet through the Weichselian. Quaternary Science Reviews 20: 591-625

Clark CD. 1993. Mega-scale glacial lineations and cross-cutting ice-flow landforms. Earth Surface Processes and Landforms 18: 1-29

Clark CD. 1997. Reconstructing the evolutionary dynamics of palaeo-ice sheets using multi-temporal evidence, remote sensing and GIS. Quaternary Science Reviews 16: 1067-1092

Clark CD. 1999. Glaciodynamic context of subglacial bedform generation and preservation. Annals of Glaciology 28: 23-32

Clark CD, Hughes ALC, Greenwood SL, Spagnolo M, Ng FSL. 2009. Size and shape characteristics of drumlins, derived from a large sample and assoicated scaling laws. Quaternary Science Reviews 28: 677-692

Clark CD, Meehan RT. 2001. Subglacial bedform geomorphology of the Irish Ice Sheet reveals major configuration changes during growth and decay. Journal of Quaternary Science 16: 483-496. DOI: 10.1002/jqs.627

Close MH. 1867. Notes on the general glaciation of Ireland. Journal of the Royal Geographical Society of London 1: 207-242

Close MH. 1867. Notes on the general glaciation of the Rocks in the Neighbourhood of Dublin. Journal of the Royal Geological Society of Ireland 11: 3-12

Congalton R, Green K. 1999. Assessing the accuracy of remotely sensed data: Principles and practices. CRC/Lewis Press

Dunlop P, Clark CD. 2006. Distribution of Ribbed Moraine in the Lac Naococane Region, Central Québec, Canada. Journal of Maps 2: 59-70

Greenwood SL, Clark CD. 2009. Reconstructing the last Irish Ice Sheet 2: A geomorphologically-driven model of ice sheet growth, retreat and dynamics. Quaternary Science Reviews 28: 3101-3123

Hubbard A, Bradwell T, Golledge N, Hall A, Patton H, Sugden D, Cooper R, Stoker M. 2009. Dynamic cycles, ice streams and their impact on the extent, chronology and deglaciation of the British–Irish ice sheet. Quaternary Science Reviews 28: 758-776

Hughes ALC, Clark CD, C.J. J. 2010. Subglacial bedfordms of the last British Ice Sheet. Journal of Maps 6: 543-563

Kleman J, Borgström I. 1996. Reconstruction of palaeo-ice sheets: the use of geomorphological data. Earth Surface Processes and Landforms 21: 893-909

Kleman J, Hättestrand C, Stroeven AP, Jansson KN, De Angelis H, Borgström I. 2006. Reconstruction of paolaeo-ice sheets - inversion of their glacial geomorphological record. In Glaciology and Earth's Changing Environment, Knight P (ed). Blackwell; 192-198.

Margold M, Jansson KN, Kleman J, Stroeven AP. 2011. Glacial meltwater landforms of central British Columbia. Journal of Maps 7: 486-506

Punkari M. 1980. The ice lobes of the Scandinavian ice sheet during the deglaciation of Finland. Boreas 9: 307-310

Rose J. 1989. Glacial stress patterns and sediment transfer associated with the formation of superimposed flutes. Sedimentary Geology 62: 151-176

Rose J, Letzer J. 1977. Superimposed drumlins. Journal of Glaciology 18: 471-480

Rose J, Smith MJ. 2008. Glacial geomorphological maps of the Glasgow region, western central Scotland. Journal of Maps v2008: 399-416

Schowengerdt RA. 2006. Remote Sensing: Models and Methods for Image Processing. Academic Press
Smith MJ. 2003. Technical Developments for the Geomorphological Reconstruction of Palaeo Ice Sheets from Remotely Sensed Data, Unpublished PhD Thesis, Department of Geography. University of Sheffield.
Smith MJ. 2011. Digital Mapping: visualisation, interpretation and quantification of landforms. In Geomorphological Mapping: methods and applications, Smith MJ, Paron P, Griffiths J (eds). Elsevier: London; 225-251.
Smith MJ, Clark CD. 2005. Methods for the visualisation of digital elevation models for landform mapping. Earth Surface Processes and Landforms 30: 885-900
Smith MJ, Knight J. 2011. Palaeoglaciology of the Last Irish Ice Sheet Reconstructed from Striae Evidence. Quaternary Science Reviews 30: 147-160
Stokes CR. 2000. The geomorphology of palaeo-ice streams: identification, characterisation and implications for ice stream functioning, Unpublished PhD Thesis, Department of Geography. University of Sheffield.
Stokes CR, Clark CD. 2003. The Dubawnt Lake palaeo-ice stream: evidence for dynamic ice sheet behaviour on the Canadian Shield and insights regarding the controls on ice stream location and vigour. Boreas 32:
Stokes CR, Clark CD, Storrar R. 2009. Major changes in ice stream dynamics during deglaciation of the north-western margin of the Laurentide Ice Sheet. Quaternary Science Reviews 28: 721-738
Stokes CR, Spagnolo M, Clark CD, O Cofaigh C, Lian OB, Dunstone RB. 2013. Formation of mega-scale glacial lineations on the Dubawnt Lake Ice Stream bed: 1. size, shape and space from a large remote sensing dataset. Quaternary Science Reviews 77: 190-209
Storrar R, Stokes CR. 2007. A Glacial Geomorphological Map of Victoria Island, Canadian Arctic. Journal of Maps 3: 191-210
Sugden D, John BS. 1976. Glaciers and Landscape. Edward Arnold: London
Wright WB. 1912. The drumlin topography of south Donegal. Geological Magazine 9: 153-159
Wright WB. 1937. The Quaternary Ice Age. Macmillan, London, 478pp.

Table 1: User defined thresholds defined prior to running CLustre.

Input	**Definition**
Distance (d)	the maximum distance [m] allowed between the seed and target lineaments, based upon lineament centroids
Orientation (α)	the maximum orientation difference [degrees] allowed between the seed and target lineaments
Length (l)	maximum difference in length [m] allowed between seed and target lineaments
Seed ID (s)	the ID of the initial seed lineament from which the clustering begins

Table 2. Aggregated results from 100 simulations for each of the 15 synthetic datasets (see Figure 5). Mean number of lineaments classified for flow set 1/flow set 2 and unclassified. Overall accuracy is shown by the results for flow set 1. User accuracy takes in to account the misclassification of flow set 2 lineaments.

Synthetic Dataset	**Flow Set 1 (%)**	**Flow Set 2 (%)**	**Unclassified (%)**	**User Accuracy (%)**
5a	99.8	0.0	0.2	100.0
5b	99.6	0.0	0.4	100.0
5c	99.8	98.8	0.2	50.2
5d	99.8	0.0	0.2	100.0
5e	98.9	0.0	1.1	100.0
5f	99.9	0.0	0.1	100.0
5g	99.9	0.0	0.1	100.0
5h	99.9	99.0	0.1	50.2
5i	99.8	0.0	0.2	100.0
5j	99.9	0.0	0.1	100.0
5k	92.2	0.0	7.8	100.0
5l	89.1	44.3	10.9	66.8
5m	97.3	94.0	2.7	50.9
5n	93.5	0.0	6.5	100.0
5o	94.9	0.0	5.1	100.0

Table 3. Summary data of flow patterns classified for the Dubawnt Case Study (n=11,825). For each for the eight flow patterns (Figure 9), the seed lineament and thresholds, along with summary statistics for lineament length and density, are presented. Comparative results from Stokes (2000) are highlighted in grey.

	User Thresholds				Summary Statistics (Length)					
Flow Pattern	**Seed ID**	**Dist (km)**	**Orient (°)**	**Len (km)**	**n**	**Mean (km)**	**StDev (km)**	**Min (km)**	**Max (km)**	**Density (lins/km^2)**
1	5175	10	2	3	8407	1.9	1.2	0.3	12.7	0.11
(blank)					8856	1.8			12.7	
2	775	15	4	5	103	3.1	1.4	0.9	6.9	0.11
(7)					120	3.0			9.1	
3	735	15	4	5	136	1.9	0.7	0.7	4.5	0.07
(6)					232	1.7			4.5	
4	1729	10	10	3	129	1.7	0.8	0.4	4.9	0.07
(3)					140	1.6			4.8	
5	857	20	3	3	555	0.4	0.2	0.1	1.4	0.06
(4)					812	0.4			1.4	
6	2386	35	5	3	1015	0.6	0.3	0.2	2.2	0.02
(1,10)					1053	0.6,1.1			2.2	
7	1261	20	3	3	109	0.9	0.3	0.3	1.9	0.08
(12)					131	0.9			1.9	
8	1253	35	3	3	86	0.9	0.2	0.4	1.8	0.03
(11)					102	3.9			1.8	
Unclassified					1,285	1.1	1.2	0.2	11.5	-

Table 4. Summary data of flow patterns classified for the Victoria Island Case Study (n=54,468). For each for the 58 flow patterns, the seed lineament and thresholds, along with summary statistics for lineament length and density, are presented. Each flow pattern is numbered on Figure 10.

	User Thresholds				Summary Statistics (Length)						
Flow Pattern	**Seed ID**	**Dist (km)**	**Orient (°)**	**Length (km)**	**n**	**Mean (km)**	**StDev (km)**	**Min (km)**	**Max (km)**	**Density (lins/km^2)**	**Stokes' Flow Pattern**
1	7958	10	2	3	6140	0.01	6.39	1.21	0.76	0.39	16
2	10301	10	2	3	4187	0.02	6.29	1.14	0.72	0.30	29,39,42,43
3	43083	10	2	3	7833	0.01	7.26	0.95	0.65	0.54	53,54,55
4	46382	10	2	3	377	0.15	2.72	0.83	0.41	0.68	60
5	27546	10	2	3	3261	0.04	6.38	0.91	0.66	0.33	20
6	13761	10	2	3	3966	0.03	5.36	1.06	0.62	0.33	31
7	14572	10	2	3	4272	0.01	6.06	0.93	0.65	0.36	17,27
8	17724	10	2	3	2992	0.01	3.07	0.6	0.31	0.52	33,34,35,45,46,47
9	21909	10	2	3	1357	0.17	5.18	1.5	0.8	0.18	19
10	33484	10	2	3	2491	0.01	6.67	0.81	0.58	0.40	17
11	9762	10	2	3	446	0.02	3.39	1.17	0.52	0.70	37
12	8949	10	2	3	300	0.21	2.66	0.94	0.41	1.33	24
14	50571	10	3	3	1034	0.11	4.3	0.91	0.54	0.62	76
15	49682	10	3	3	853	0.19	2.92	0.69	0.35	0.57	80
16	49400	20	2	3	590	0.2	2	0.71	0.31	0.23	78
17	15469	15	2	3	788	0.02	4.53	0.95	0.47	0.23	30
18	47584	12	3	2	548	0.01	4.9	0.87	0.52	0.53	56
19	29887	12	3	3	571	0.16	3.07	0.63	0.37	0.43	61,63
20	48253	12	3	3	718	0.18	2.39	0.67	0.32	0.48	74,75
21	52340	12	3	3	958	0.01	2.98	0.54	0.27	0.68	85
22	32074	15	4	5	817	0.19	10.22	1.27	1.13	0.15	66,67,68,69
23	19393	10	2	3	360	0.19	1.85	0.66	0.27	0.35	48
24	53922	15	3	3	266	0.28	4.52	1.45	0.75	0.19	17
25	49126	15	3	3	276	0.25	3.47	0.85	0.46	0.28	
26	47087	10	2	3	303	0.25	3.01	1.12	0.55	0.26	58

27	47970	10	2	3	52	0.01	0.8	0.44	0.14	0.61	59
28	41317	25	5	5	351	0.23	2.52	0.89	0.37	0.10	73
29	41299	10	3	4	75	0.25	2.46	0.85	0.42	0.26	71,73
30	51639	20	4	4	113	0.35	3.42	1.05	0.47	0.08	80
31	1813	15	4	4	210	0.41	4.61	1.4	0.65	0.35	16
32	18903	10	4	7	176	0.35	10.9	2.34	1.68	0.09	16
33	22193	10	3	3	68	0.39	2.29	1.1	0.47	0.30	16
34	4257	10	3	3	785	0.01	2.21	0.41	0.26	1.15	40
35	4843	10	5	5	164	0.03	4.57	0.93	0.61	0.26	18
36	38161	10	4	8	223	0.26	3.86	1.11	0.67	0.40	70
37	37204	10	4	5	135	0.15	1.65	0.65	0.28	0.47	70
38	31273	10	3	3	766	0.09	2.6	0.55	0.32	0.50	52,62
39	46100	10	3	3	270	0.21	2.28	0.78	0.38	0.76	60
40	41190	15	5	5	41	0.22	1.68	0.61	0.28	0.20	78
41	47267	15	5	5	80	0.3	2.94	1.02	0.42	0.29	58
42	48982	15	5	5	133	0.22	1.45	0.6	0.22	0.95	57
43	51381	15	3	5	96	0.2	1.41	0.56	0.26	1.12	86
44	51503	15	3	5	81	0.16	1.58	0.73	0.31	0.86	87
45	14551	15	3	3	72	0.3	1.67	0.71	0.22	0.17	
46	50162	15	3	3	74	0.37	1.88	0.92	0.37	0.24	81
47	53352	10	4	4	88	0.22	1.63	0.64	0.28	0.54	80
48	52851	10	3	4	137	0.15	1.44	0.54	0.25	0.66	82
49	54251	10	4	4	50	0.38	2.81	1.18	0.47	0.25	17
50	8671	10	4	4	76	0.01	2.68	1.07	0.57	0.27	26
51	9015	10	4	4	26	0.02	1.37	0.74	0.31	0.92	26
52	9022	10	4	4	20	0.39	1.69	0.84	0.29	0.38	26
53	19195	10	4	4	237	0.18	5.41	0.94	0.86	0.22	35
54	40406	15	3	3	129	0.24	3.36	1.04	0.52	0.15	31
55	13213	15	3	3	112	0.45	4.14	1.36	0.71	0.22	28
56	26083	15	3	3	687	0.14	5.26	0.9	0.72	0.12	20,49,50
57	22932	20	3	3	121	0.13	1.03	0.4	0.16	0.45	44
58	26135	20	3	4	78	0.26	7.15	1.31	1.25	0.13	51

Figure 1: Data collection involved in glaciological inversion modelling: (a) oblique view of a lineament (Photo: J. Rose), (b) base satellite imagery, (c) manual mapping of individual landforms, (d) qualitative data reduction to form flow patterns and (e) flow sets (see Smith, 2003).

Figure 2: (a) shows individual mapped lineaments, whilst (b) and (c) show two alternative interpretations of their formation. Do they represent a transgressive retreating margin or two separate flow patterns? (d) illustrates that by grouping lineaments and reviewing their characteristics (e.g. spacing and length), there can be a basis for interpretation (after Clark,1993).

Figure 3: Two contextual scenarios with similar morphological patterns where spatial heuristics may fail to form flow patterns. (a) two flow patterns that cross-cut yet lineament orientations coincide in-part and (b) a single time-transgressive flow pattern which displays cross-cutting, low orientation conformity and abrupt morphometric changes (after Clark, 1999). Without further contextual information it is difficult to interpret the flow patterns in these two scenarios.

Figure 4: Graphical depiction of the clustering algorithm. (i) seed lineament selected (random or identified), (ii) nearest lineaments located, (iii) cluster membership evaluated based upon length/orientation, (iv) two furthest lineaments become secondary seed lineaments, (v) nearest unallocated lineaments are selected and (vi) cluster membership is evaluated. The algorithm iterates until no more lineaments can be selected. At (1) the maximum distance to a lineament is exceeded and so cannot be selected. At (2), the maximum deviation in orientation is exceeded and so cannot be selected.

Figure 5 Synthetic datasets generated for testing with CLustre. a,f,k show the simplest scenarios with one flow set that is straight, curving and divergent. b,g,l and c,h,m show end-members for a more complex two flow set example with straight, curving and divergent lineaments orthogonal and parallel to one another. d,j,n and e,j,o show the same datasets, but with the extra variable of different lineament lengths for each flow set.

Figure 6. Results of classification simulations using the 15 different synthetic datasets. Each dataset contains 100 lineaments per flow set, with the simulation iteratively run using each lineament as a seed (i.e. 100 classifications per dataset). The histograms show the frequency of FS1 (grey bar) and FS2 (black bar) lineaments classified based upon the frequency of simulations. For example (b) shows all FS1 lineaments are correctly classified, whilst all FS2 lineaments remain correctly unclassified.

Figure 7. Procedure for the extraction of flow patters for the Dubawnt Case Study. (a) Flow pattern 1; (b) Flow pattern 2 with initial thresholds; (c) Flow pattern 2 with modified thresholds; (d) Flow pattern 6 with initial thresholds and (e) Flow pattern 6 with modified thresholds (see Table 3). Grey box indicates the extent of 7b, 7c and 7d. Grey box indicates the extent of Figure 9.

Figure 8. Histograms of length for the six flow patterns classified in the Dubawnt Case Study (0=unclassified).

Figure 9. Extract of classified flow patterns for the Dubawnt Case Study. The different flow patterns (1=red, 2=blue, 4=brown, 5=pink, 6=green; unclassified=black) illustrate the robustness of the classification procedure and complexity that can be highlighted (Projection: UTM13N, Coordinates: Metres). Note a number of unclassified lineaments that are part of flow pattern 1 but have been excluded on the basis of length.

Figure 10. Extract of classified flow patterns for the Victoria Island case study. The different flow patterns are numbered (see Table 4); black lineaments are unclassified. (Projection: UTM13N, Coordinates: Metres).